

# Analyzing Student's Academic Performance Using Multilayer Perceptron Model

Gino Sinthia, M. Balamurugan

*Abstract---* Identification of the student's behavior in the class room environment is very important. It helps the lecturer to identify the needs of the students. It also aids in identifying the strength and weakness of the individual and guide them to improve on their performance. Observing and supervising the students regularly can improve their performance. The data has been collected from 120 students who took the common the course taught by two different lectures. The students were observed based on the internal assignments and quizzes and the model exam given by the respective lecturers. In this paper the students are categorized into different groups based on their performance using Multilayer Perceptron (MLP) and also different factors which are influencing the performance of the students are identified.

*Keywords---* Student's Performance, Machine Learning, Multilayer Perceptron, k- Nearest Neighbor (KNN), K-means.

## I. INTRODUCTION

Classroom observation is very important and it has been used widely in the past decades. Automated systems are being used to observe the behavior of the students in the class which are very easy to observe, compared to the traditional ones. If observations are done by human, is difficult to implement when there are many people. Observation of student's performance in the classroom helps us to find their behavior pattern and also help us to provide proper guidance and also helps the students to improve their performance. Many scholars emphasized that the participation in the classroom environment improves the performance of the students. One of the key indicator to evaluate the level and quality of the education is academic performance. It also plays vital role of higher education graduation which motivate the students towards their goals. So performance analysis is very important task and also one of the difficult task too. Since the factors that affect the performance is diverse and complex because everyone is different and each one has different way of doing things. Everyone follows different behavioral pattern the behavioral pattern of the student changes from time. [1]The learning effectiveness and the performance may not follow the same pattern throughout even in the same learning environment due to various reason and finding that reason is one of the main tasks.

Many students from engineering colleges either quit or fail due to their poor performance. They might not be placed after graduation. So identifying the students who are in the verge of failing and helping them can improve their performance and also save them from quitting. Performance of the students can be observed by continuous internal

assessments. Based on these assessments the final score can be predicted beforehand. It is observed that the students who participates during the class activities are performing well during the internal exam and outsmarts others in the final exam.

Many data mining techniques are available such as, classification, clustering, Regression and so on.

In this paper we characterize the student's performance based on their academic contribution and survey which has been conducted among the students. Here we have compared the performance of KNN, K-Means and MLP.

## II. LITERATURE SURVEY

In the paper [2] purposed by Shadi Esnaashriors, et al, it had been analyzed with second year students from Business School, New Zealand university. The data was taken from a course which lasted for 12 weeks for 103 students. The analysis is done based on the student taking the online video lecture and attending the quiz after every module. The participation is based on the tournament conducted by the lecturers. Students were grouped based the video lectures on their performance in the quizzes. K-means Clustering algorithm was used to group the students. Based on the analysis four clusters (k=4) were formed to analyses the students behavior.

Xi Zhang et al. [3] mentioned in their work that the poor performance of the college students is identified based on the behavior pattern. The dataset consist of student performance in 12 courses for the duration of 18 weeks for 302 student and it has 24 relevant features. The features includes the time spending for studying and other activities such as spending time in campus, library, and other personal work. The user behavior pattern is recorded from the smart card and the features that are required are extracted. The multi-model is used for the classification and the recall value is high enough to discover the students with poor performance.

The work provided by Liansheng Jia et al [4] analyze the behavior of the student based on the Chinese reading assessment. The dataset had been collected from 349 students from 4<sup>th</sup> grade in Hubei province of China. Here K-means is used for clustering the students based on their performance. Three clusters were formed. The students spent less time in reading the article and answering question is cluster 1 but they lacked reading ability. Cluster 3 is similar to the cluster 1 but the students have high ability and took the test more seriously. But clusters 2 were the students who spent longest time in reading and answering the questions. 14% of students in cluster 1, 15% of student fall

Manuscript received February 01, 2019

Gino Sinthia, Dept. of Computer Science and Engineering CHRIST (Deemed to be University), Faculty of Engineering, Bangalore, Karnataka, India.

M. Balamurugan, Dept. of Computer Science and Engineering CHRIST (Deemed to be university), Faculty of Engineering, Bangalore, Karnataka, India.

under cluster and 25 % of students in cluster 3.

The paper written by ferbrianti Widyahastuti et al [5] predicts the performance of the students in the final exam. Here Linear Regression and multi-layer perceptron model are used for prediction and comparing the performance of both the models based on the prediction performance, error rate and accuracy. The data is collected from 50 undergraduate students with the attributes like posting in the discussion group and attendance is taken in to account for classification. Both the algorithms are compared using the Absolute Error and the prediction result shown. The error rate is 12.63 and 11.85 in Linear Regression and Multilayer Perceptron respectively. It shows that the multilayer perceptron model has good prediction accuracy compared to linear regression.

The work purposed by Tao Zhang et al [6] predicts the fluctuations in the performance of the students based on the short –term and long-term data. The long term data for example students name student's ID doesn't change. This model predicted 1995 student's ranking for the next exam and compared the outcome with the actual academic results. Comparative study of Step regression, Decision tree, Logistic regression and SVM regression made based on the prediction accuracy. The results are shown that the SVM regression (prediction 77%) outsmart than other algorithms. The main aim of this paper [7] is to analyze and monitor the students' performance by using clustering and deterministic model. K-Means clustering is more efficient when it comes in getting the results in short period of time. Here 65 students' data has been collected for one particular semester. The algorithm is considered for different values of k where k=3, 4, 5. For each cluster the cluster size and performance is determined.

The work carried out in [8] is based on the eagerness to work, programming practice to analyze students' performance for the introductory programming course. The information is collected automatically from 152 students with around 6600 snapshots in six week programs. Three groups are considered based on their performance Fail, Pass, Excellent. They have used Bayesian network to classify the students into the groups. The accuracy is estimated with Confusion matrix. Initial prediction accuracy in end of the second week is 65% and by end of the course the accuracy is 78%.

This paper [9] predicts the performance of the engineering students. Principle Component Analysis (PCA) is used for analyzing and also for establishing the early warning. Here the students were categorized into Successful(S), Probation (P), Failed (F), Not failed but not received(I).90 records were collected out of which 29 =S, 21=P, 30=F,10=I.

In paper [10], the proposed system predicts the performance of the students with the generated rules by using different data mining techniques. Here classification techniques are used to classify the students based on their grades. The prediction performance are 58.3%, 59.1%, 61.4%, 45.7% 58.3% under J48, Simple cart, BFTree , Random Tree and J48 graft respectively. This system helps the lecturers to identify the students who are in the verge of failing.

In this paper [11], the relation between the students' performance in the mathematics and their choice of course is analyzed. Different statistics techniques such as Paired t-test, Anova, Two way Anova, sample paired t test are used to find the relation. Based on the significance value (p value), it is concluded that there is no significance between the performance and the choice of the course.

### III. SYUDY DESIGN

Participants in this study are the students of faculty of engineering at CHRIST (Deemed to be University) for the one department core subject. This course lasted for 6 months. Data is available from 120 students who took the course in their second year of engineering. During the course, students were assessed by Continuous Internal Assessments (CIA). CIA1 and CIA3 which were the collection of assignments, quizzes and class test and CIA2 is conducted as model exam. A survey was conducted to get the feedback from the students about course and other factors.

The dataset includes the feedback about the class conduction, toughness of the course, time spent for the studying the particular subject (daily, weekly or before exam). Since it's considered in the classroom environment their travel time is taken into consideration and living status (with parents or guardian or others) too. All these personal information are directly collected from the students through feedback.

The data such as CIA1, CIA2, and CIA3 are directly collected from lecturers. CIAs are divided into different components like assignments, Quizzes and class test. These components are analyzed by the lecturers. Based on the data collected from the students and assessment report of CIA components, the performance of the students has been analyzed and most important factors which are influencing their academic performance are identified. Table 1 shows all the attributes which we considered as influencing factors of student's performance.

### IV. METHODOLOGY

In order to characterize the student's behavior based on their participation and to identify the factors that influence their performance, machine learning algorithm can be used.

Here, we have used MLP which is a deep artificial neural network algorithm. It has an input layer to receive the input and the output layer to make the decision about the input. The computational engine of MLP consist of hidden layers which are capable of approximating any type of continuous function.

It is one of the supervised learning problem. The model is trained with the set of input and output, then the correlation or the dependencies are learnt from the given input. In order to reduce the error, bias function is required to be adjusted with the help of back propagation Root Mean Square Error (RMSE) is used to for measuring the error.



**Table 1: List of Attributes in the feedback form**

S.No	Attributes
1	Course taught by
2	Feedback on class conduction
3	Toughness of the Course
4	Family Size
5	Living with
6	Mother's Education
7	Father's Education
8	Mother Working
9	Father Working
10	Travel Distance
11	Study Time
12	Failures
13	Discussion with Faculty
14	Participated in Extra-Curricular activities in college
15	Attendance percentage for particular subject

are study time, Travel time, Toughness of the course, Activities, Discussion, attendance.

Other than these above mentioned factors, remaining factors dint have much influence on the student's performance.

CIA 2 is one of the main factors which help to predict the performance of the students since its significance is less. From the figure 1, it is understood that the students who scored above 40 in the CIA 2, they performed well in the final exam and they falls in the class "good". On the other hand students who score very less score in the CIA 2 tends to perform badly in final exam and they falls in the class "fail". So this can be taken as one of the critical factor for the students' performance analysis.

**V. DATA ANALYSIS AND RESULTS**

The data are collected through feedback survey and academic results are preprocessed for the analysis. In our dataset, the marks of CIA1 and CIA3 are taken for 20 and CIA 2 is considered out of 50.

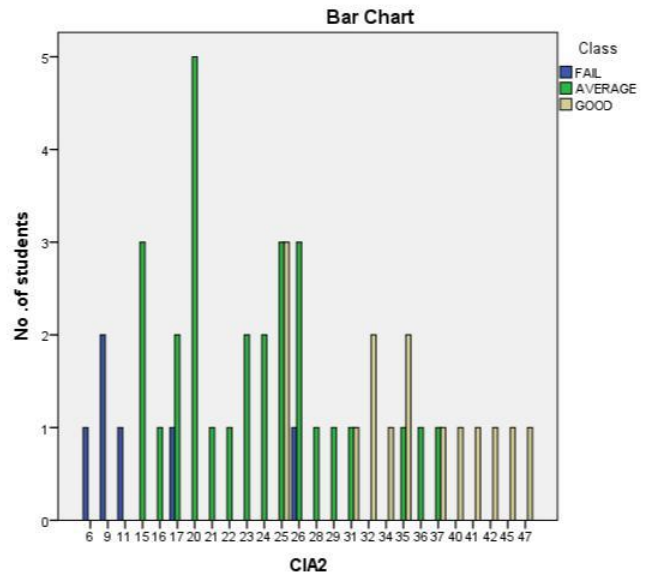
Multilayer Perceptron model is applied to the dataset and the students were segregated in 3 groups as: Good, Average and Fail. Here we need to identify the factors which are influencing the students' performance. Chi-square test is performed to find the significance value of each attribute.

The overall significance values which affect students' performance have been given in table 2.

**Table 2: Significance value**

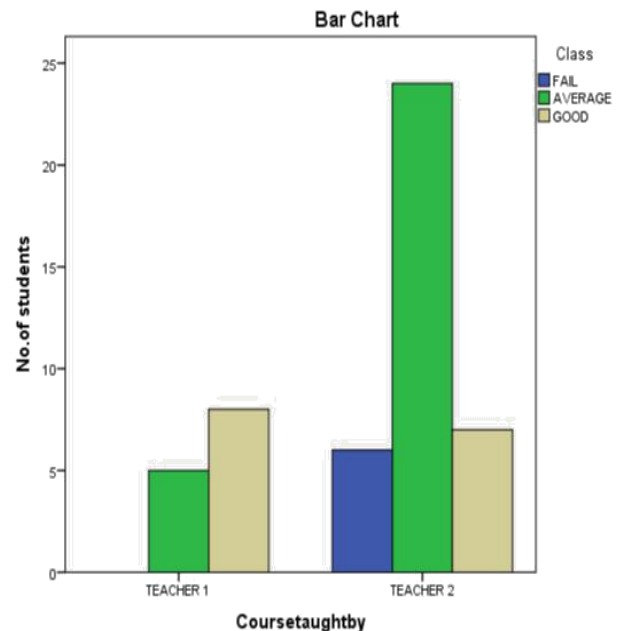
Attributes	Significance Value
CIA 1	0.01
CIA 2	0.00
CIA 3	0.03
Toughness of the course	0.054
Study Time	0.05
Course taught by	0.028
Living With	0.063
Activities	0.603
Attendance	0.44
Travel Distance	0.15
failures	0.448

Based on the significance values, it is observed that CIA 1, CIA 2, CIA 3 are influencing the students majorly. Out of which the CIA 2 influence is more in the performance. Others factors that also has the influence on the performance.



**Figure 1: Performance in CIA 2 Vs Final exam**

From figure 2 it can be interpreted that the no student's falls under the fail category when the course taught by teacher 1 and the number of the students who falls under the good category is more that of the course taught by the teacher 2. It is observed that the teacher's role also plays a one of the important role in student's performance in that respective course.



**Figure 2: Role of Teacher**

From Figure 3, it is identified that the students who falls under the class "Good" have given good feedback about the class conduction.



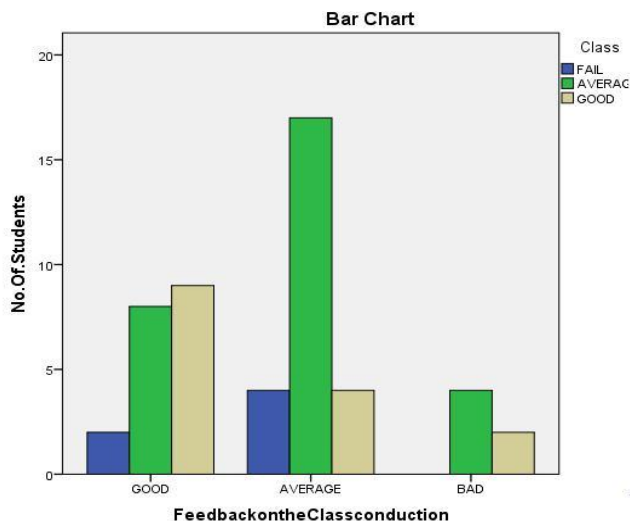


Figure 3: Performance of Students vs. Class Conduction

In figure 4, Based on their study time it is understood maximum number of students study before the exam. So mainly students falls under the average category

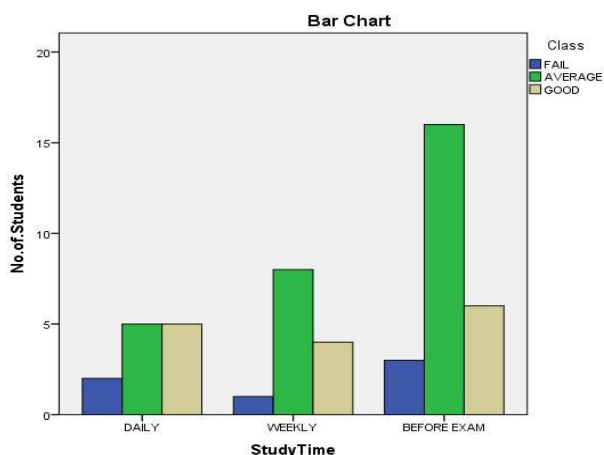


Figure 4: Performance of Students vs. Study Time

From figure 5, despite of the toughness of the course, the student's performance is average even though they felt the course is manageable or difficult during the feedback survey.

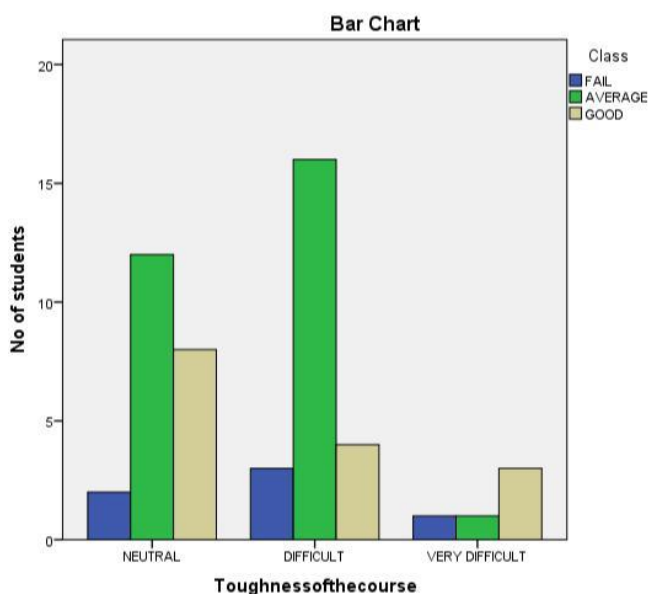


Figure 5: Performance vs. Toughness of the Course

Here we have used 50 unlabeled records for the prediction. The predicted results are compared with two more algorithms KNN and K-Means. The following are the results of each algorithm's performance.

Table 3: KNN

	Average	Good	Fail	Prediction Accuracy
Average	19	5	4	67.8%
Good	3	12	0	80%
Fail	3	0	4	57.1%

Table 3 shows the results of KNN with the value of k = 3 its. Out of 28 records 19 records were predicted correctly as Average. 12 records were predicted as Good out of 15 records. 4 records were predicted correctly as fail out of 7.

Table 4: K-Means

	Average	Good	Fail	Prediction Accuracy
Average	20	4	4	71.4%
Good	1	14	0	93.3%
Fail	2	0	5	71.4%

Table 4 shows the results of K-Means algorithm where k=3. 20 records were predicted as Average, 14 records predicted as Good and 5 records were predicted correctly.

Table 5: MLP

	Average	Good	Fail	Prediction Accuracy
Average	22	3	3	78.6%
Good	0	15	0	100%
Fail	1	0	6	85.7%

Table 5 represents the results of MLP, 22 records 78.6% is predicted correctly as average, all the 15 100% records were predicted correctly as Good and 6 records 85.7% is predicted and Fail.

Table 6: Performance Comparison

Algorithm	Performance
KNN	70%
K-means	78%
MLP	86%

From the table 6 it's understood that the accuracy of MLP is high compared to the KNN = 70% and K-means = 78% and MLP with highest accuracy of 86%.

## VI. CONCLUSION

According to our analysis, it has been observed that performance in CIA 2, Study time, the role of faculty in teaching, Toughness of the course are highly influencing student's performance in their academics. Multi-layer Perceptron(MLP) is working good for analyzing the student's performance in considered environment and dataset, compared with other algorithm KNN and K-Means.





This analysis can be used to identify the students who are in verge of failing and counsel them to improve their performance.

15. Institute, Universiti Kuala Lumpur”, International Conference on Statistics in Science, Business and Engineering,2012.

### FUTURE WORK

In future this work can be extended by considering the details of student’s participation in the class activities and their classroom behavior. Meanwhile this performance analysis work could be implemented with different machine learning algorithms.

### REFERENCES

1. C. Brooks, G. Erickson, J. Greer, and C. Gutwin, "Modelling and quantifying the behaviours of students in lecture capture environments," *Computers & Education*, vol. 75, pp. 282-292, 2014.K. Elissa,
2. Liansheng Jia,Sannyuya Liu,Yangjun Chen, Hercy N.HCheng,Wang-ChenChang,JianwenSun,” Integrating clustering and sequential analysis to explore students’ behaviors in an online Chinese reading assessment system”, 6th IIAI International Congress on Advanced Applied Informatics,2017
3. Febrianti Widyahastuti , Viany Utami Tjhin,” Predicting Students Performance in Final Examination using Linear Regression and Multilayer Perceptron”,10th International Conference on Human System Interactions (HSD), 2017
4. Tao Zhang, Yihua Xu,Peng Qi,Xin Li,Guoping Hu,” Predicting the Performance Fluctuation of Students Based on the Long-term and Short-term Data”, The Sixth International Conference of Educational Innovation through Technology,2017M. Young, *The Technical Writer’s Handbook*. Mill Valley, CA: University Science, 1989.
6. Xi Zhang, Guangzhong Sun\$,Yigong Pan, Hao Sun, Jiali Tan,” Poor Performance Discovery of College Students Based on Behavior Pattern”, IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation, 2017
7. Shadi Esnaashariors, Lesley Gardner, Michael Rehm,” Characterizing Students’ Behavior based on their Participation in Property Course in New Zealand”, 32nd International Conference on Advanced Information Networking and Applications Workshops , 2018
8. J. James Manoharan, Dr. S. Hari Ganesh, M. Lovelin Ponn Felciah, A.K. Shafreen Banu,” Discovering Students’ Academic Performance Based on GPA using K-Means Clustering Algorithm”, World Congress on Computing and Communication Technologies, 2014.
9. Arto Vihavainen, “Predicting Students’ Performance in an Introductory Programming Course using Data from Students’ownProgrammingProcess”,13<sup>th</sup> International Conference on Advanced Learning Technologies, 2013.
11. Kin Fun Li, David Rusk, Fred Song,” Predicting Student Academic Performance”, Seventh International Conference on Complex, Intelligent, and Software Intensive Systems,2013.
13. Ishwank Singh, A Sai Sabitha, Abhay Bansal,” Student Performance Analysis Using Clustering Algorithm”, 6th International Conference - Cloud System and Big Data Engineering, 2016.
14. Mazwin Tan, Halina Hassan, Nurul Na’imy Wan,” Statistical analysis on students’ performance and their preference of course in bachelor of Engineering Technology programmes in Malaysian Spanish